NOMAD
NOVEL MATERIALS DISCOVERY

MAX-PLANCK-GESELLSCHAFT

Data-driven cartography of materials properties

Luca M. Ghiringhelli
Fritz-Haber-Institut der MPG, Berlin

Workshop on Crystal Structure Prediction:
Exploring the Mendeleev Table as a Palette to Design New Materials
ICTP, Trieste, 13-18 January 2019

Data-driven materials science: the big picture

NOVEL MATERIALS DISCOVERY

- Design of new materials:
  preparation, synthesis, and characterization is complex and costly
- About 240000 *inorganic* materials are known to exist (Springer Materials)
- Basic properties determined for very few of them
- Number of possible materials: practically infinite

⇨ New materials with superior properties exist but not yet known

- Data-analytics tools will help to identify trends and anomalies in data and guide discovery of new materials

We have a dream

NOVEL MATERIALS DISCOVERY

From the **periodic table of the elements** to a **chart of materials**

Reihen	Gruppe I. — R^2O	Gruppe II. — RO	Gruppe III. — R^2O^3	Gruppe IV. RH^4 RO^2	Gruppe V. RH^3 R^2O^5	Gruppe VI. RH^2 RO^3	Gruppe VII. RH R^2O^7	Gruppe VIII. — RO^4
1	H=1							
2	Li=7	Be=9.4	B=11	C=12	N=14	O=16	F=19	
3	Na=23	Mg=24	Al=27.3	Si=28	P=31	S=32	Cl=35.5	
4	K=39	Ca=40	—=44	Ti=48	V=51	Cr=52	Mn=55	Fe=56, Co=59, Ni=59, Cu=63.
5	(Cu=63)	Zn=65	—=68	—=72	As=75	Se=78	Br=80	
6	Rb=85	Sr=87	?Yt=88	Zr=90	Nb=94	Mo=96	—=100	Ru=104, Rh=104, Pd=106, Ag=108.
7	(Ag=108)	C			22	Te=125	J=127	
8	Cs=133	Ba=137	?Di=138	?Ce=140	—	—	—	— — — —
9	(—)	—	—	—	—	—	—	
10	—	—	?Er=178	?La=180	Ta=182	W=184	—	Os=195, Ir=197, Pt=198, Au=199.
11	(Au=199)	Hg=200	Tl=204	Pb=207	Bi=208	—	—	
12	—	—	—	Th=231	—	U=240	—	— — — —

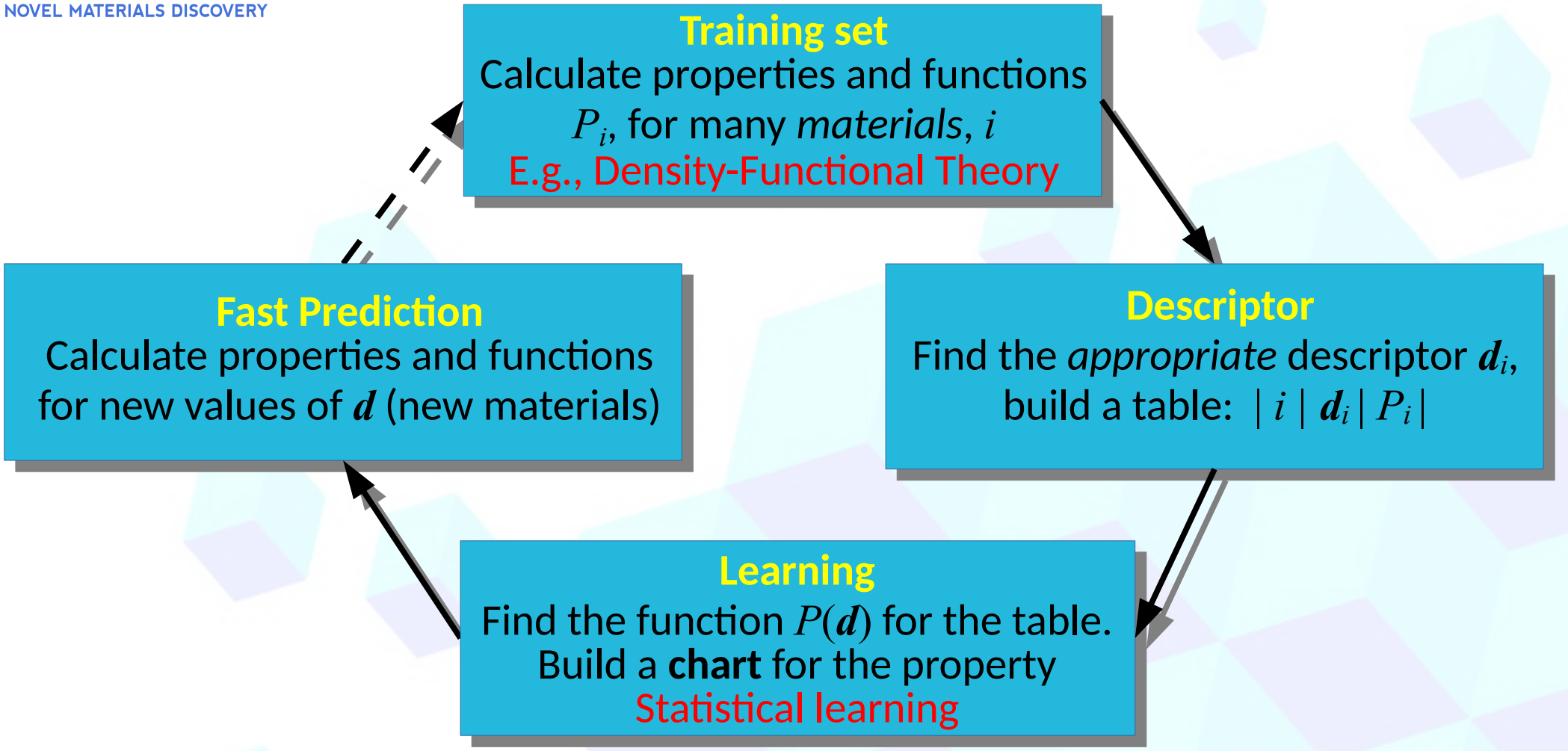
Ga=69.7 Ge=72.6

Mendeleev's **1871** periodic table

NOMAD

NOVEL MATERIALS DISCOVERY

Data analytics: an ideal flow chart



Data analytics: an ideal flow chart

NOMAD
NOVEL MATERIALS DISCOVERY

Training set
Calculate properties and functions
P_i, for many *materials*, i
ory

$\{R_\mathrm{I}, Z_\mathrm{I}\} \rightarrow$ Hamiltonian

$\{R_\mathrm{I}\} \rightarrow$ Geometry
- translational, rotational, permutational invariant
- perhaps just lattice-decoration sensitive

$\{Z_\mathrm{I}\} \rightarrow$ Chemistry

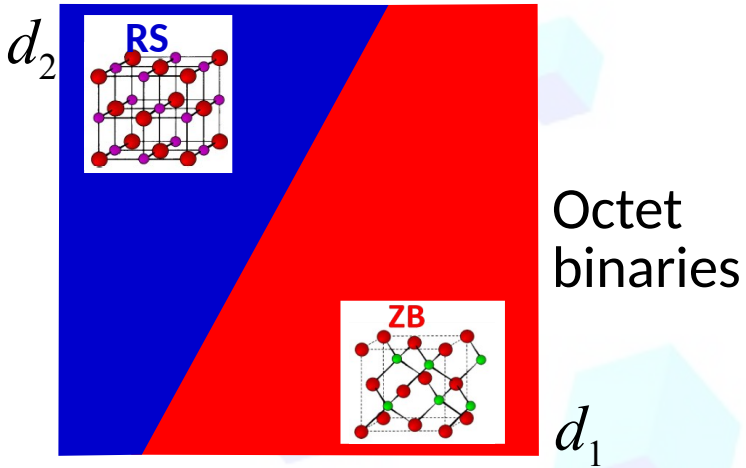
C
fo

Descriptor
the *appropriate* descriptor $\boldsymbol{d}_i$,
build a table: $| i | \boldsymbol{d}_i | P_i |$

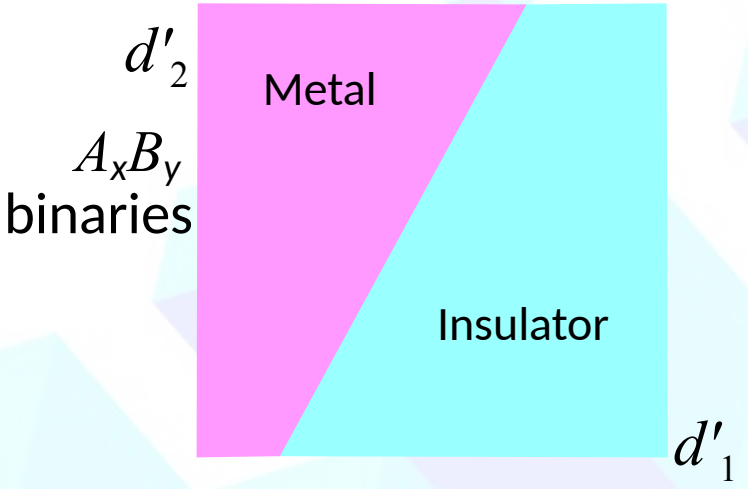
Learning
Find the function $P(\boldsymbol{d})$ for the table.
Build a **chart** for the property
Statistical learning

Learning/discovering maps of materials properties. A quantum many-body problem

NOVEL MATERIALS DISCOVERY

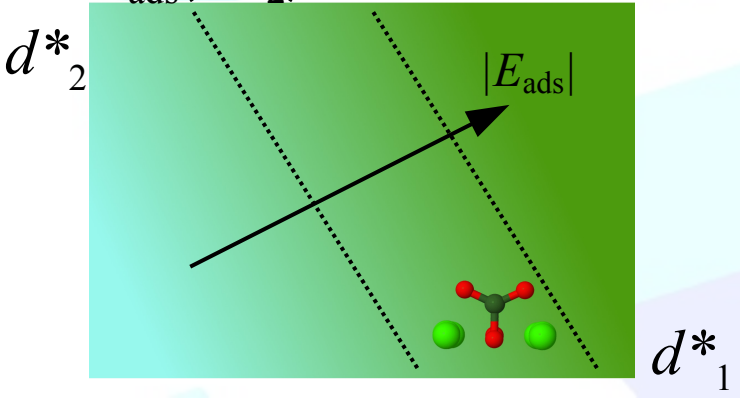


Octet binaries

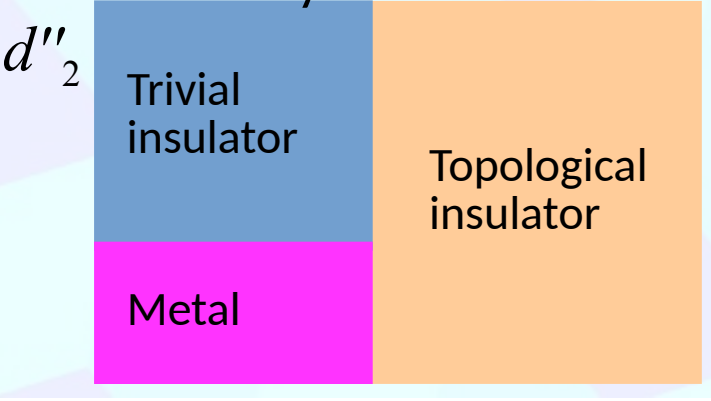


A_xB_y binaries

$E_{ads}(CO_2)$ on oxides



2D honeycomb materials



Learning/discovering maps of materials properties. A quantum many-body problem

NOVEL MATERIALS DISCOVERY

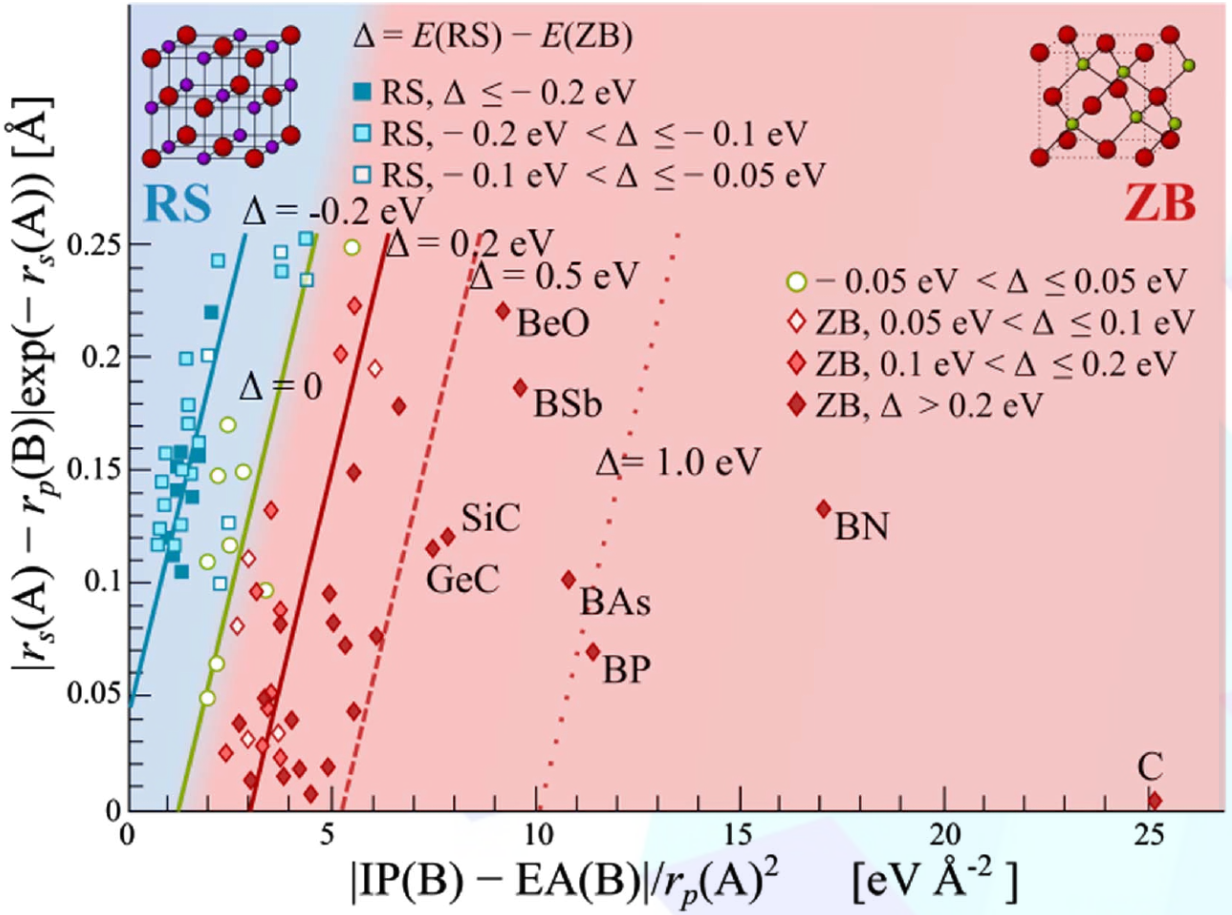
From the Hamiltonian of a physical system, in principle we can derive all properties (observables).
But in practice, the Hamiltonian is often not the starting point.

For instance, given a class of chemical compositions
(e.g., via prototype formula, such as ABX_3):

- what is the most stable crystal structure of each material in the class?
- which materials are metals / topological insulators / superconductors ?
- which material has the highest melting point?
- which materials has a surface optimal for catalysing some chemical reaction?

Compressed sensing: the quest for descriptors and predictive models

NOVEL MATERIALS DISCOVERY



Structure map with compressed-sensing algorithm, starting from 7 atomic features

LMG *et al.*, PRL 2015, NJP 2017

Data analytics with compressed sensing

Compressed sensing

Aim: finding descriptors and learning predictive models

Ansatz:

$$\boldsymbol{P} = c_1\boldsymbol{d}_1 + c_2\boldsymbol{d}_2 + \ldots + c_n\boldsymbol{d}_n$$

$\boldsymbol{P}$: property of interest

$\boldsymbol{d}_1, \ldots \boldsymbol{d}_n$: features, i.e., (nonlinear) functions of *primary features* (EA, IP, ...)

$c_1, \ldots c_n$: unknown coefficients => as few as possible are *nonzero*

$$\underset{\boldsymbol{c}\in\mathbb{R}^M}{\operatorname{argmin}} \|\boldsymbol{P} - \boldsymbol{D}\boldsymbol{c}\|_2^2 + \lambda\|\boldsymbol{c}\|_0$$

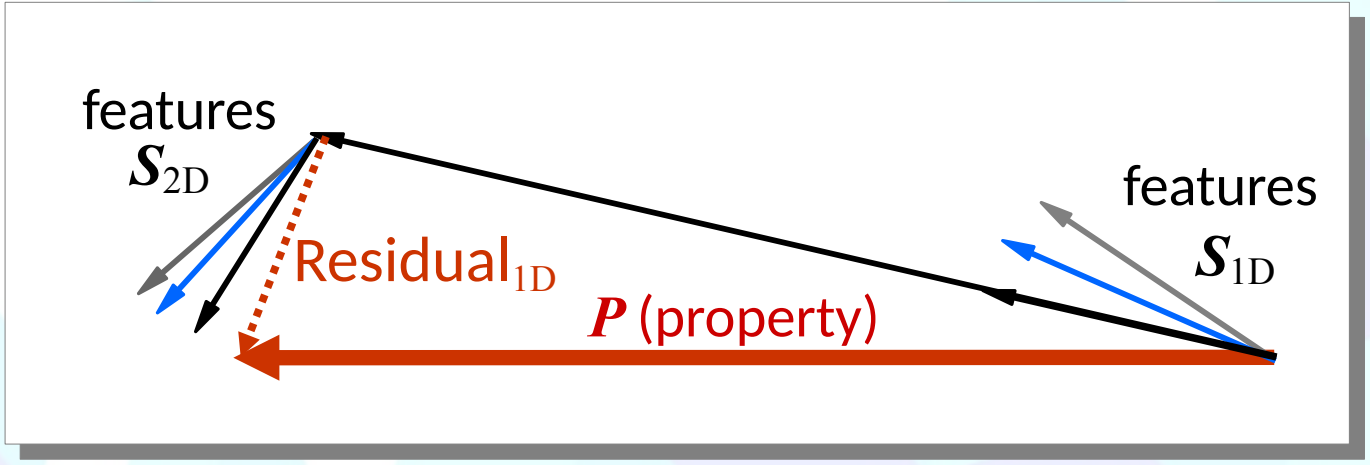
d_i: iterative construction with features and operators (+, ×, /, ², ...)

$$\frac{\mathrm{IP(B)} - \mathrm{EA(B)}}{r_p(\mathrm{A})^2}, \frac{|r_s(\mathrm{A}) - r_p(\mathrm{B})|}{\exp(r_s(\mathrm{A}))}$$

SISSO: sure independence screening plus sparsifying operator

NOVEL MATERIALS DISCOVERY

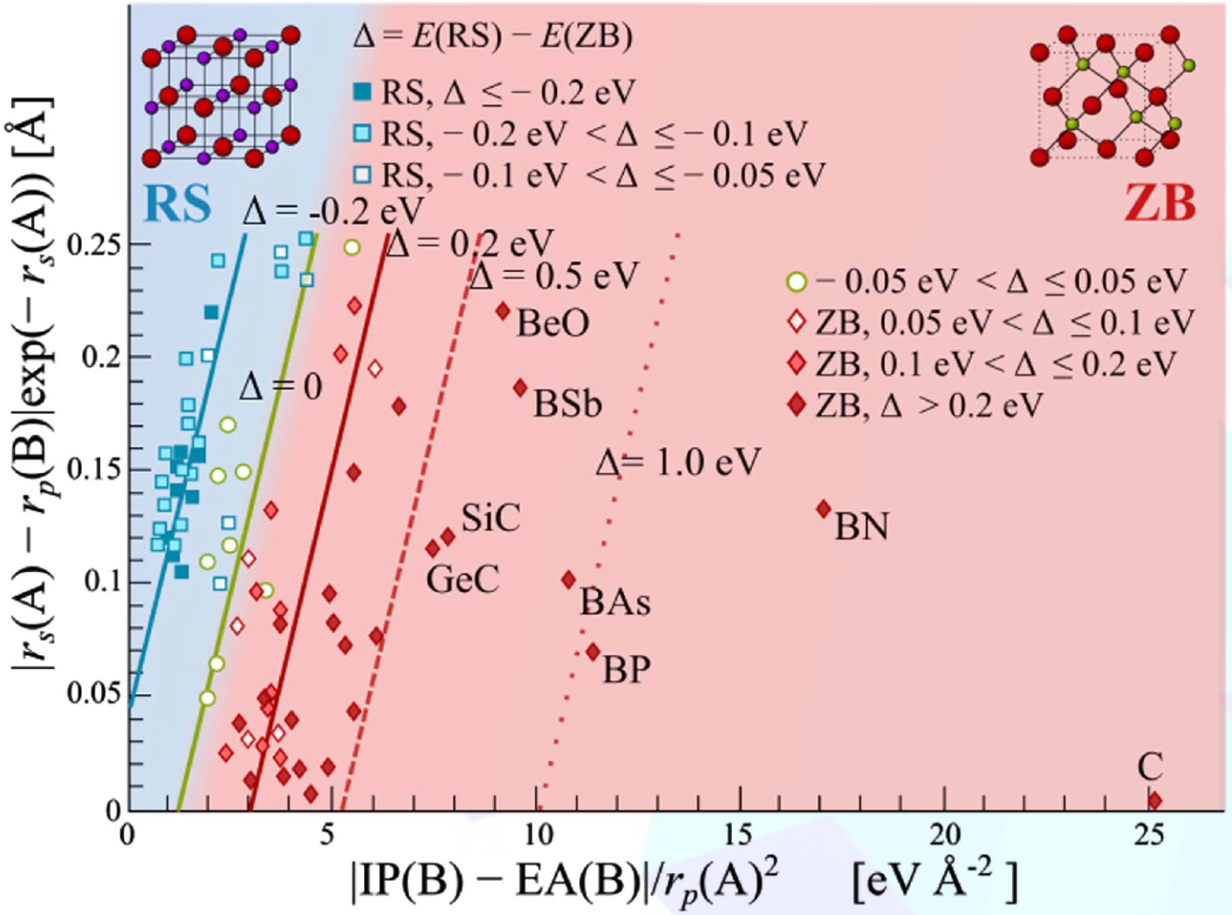
$$\underset{\boldsymbol{c}\in\mathbb{R}^M}{\operatorname{argmin}} \|\boldsymbol{P}-\boldsymbol{D}\boldsymbol{c}\|_2^2+\lambda\|\boldsymbol{c}\|_0$$



Ouyang, Curtarolo, Ahmetcik, Scheffler & LMG, PRM (2018) arXiv:1710.03319

Compressed sensing: the quest for descriptors and predictive models

NOVEL MATERIALS DISCOVERY



Structure map with compressed-sensing algorithm, starting from 7 atomic features

LMG *et al.*, PRL 2015, NJP 2017

One descriptor to rule them all: Multi-task learning

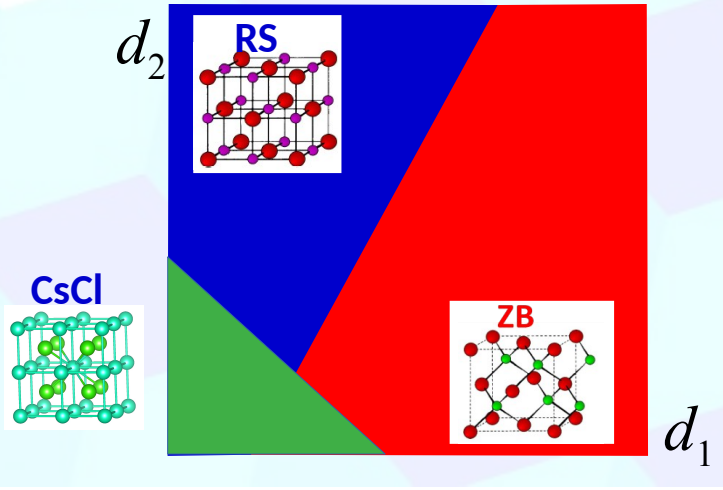
NOVEL MATERIALS DISCOVERY

$$\{P^{(1)}, P^{(2)}, \ldots, P^{N^{\mathrm{T}}}\} \longrightarrow P^k = \boldsymbol{d} \cdot \boldsymbol{c}^k$$

$$\arg\min_{\boldsymbol{c}} \left(\|\boldsymbol{P} - \boldsymbol{D}\boldsymbol{c}\|_2^2 + \lambda \|\boldsymbol{c}\|_0 \right)$$

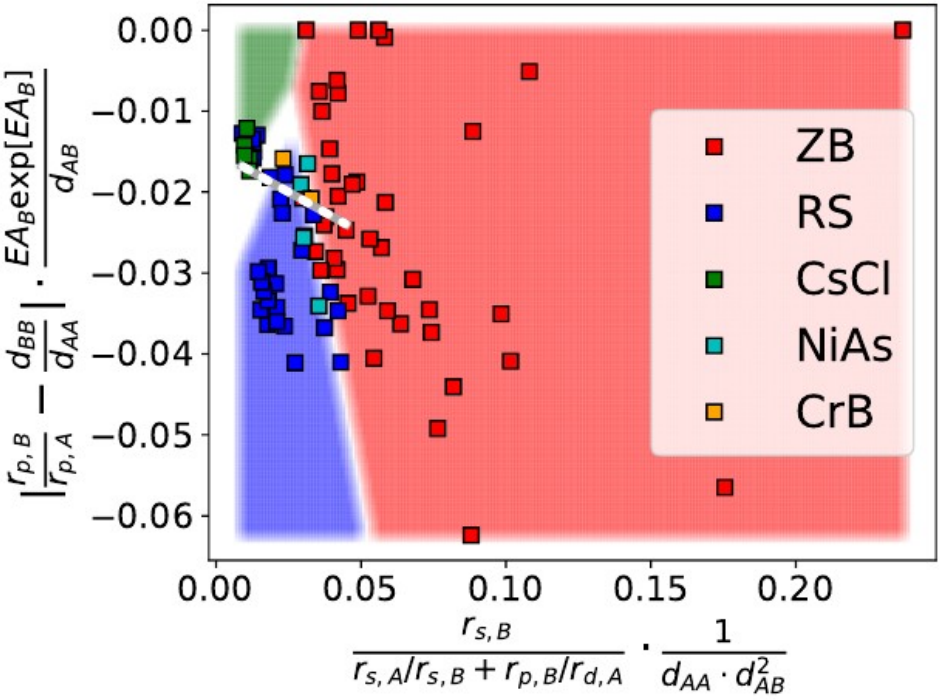
$$\arg\min_{\boldsymbol{C}} \sum_{k=1}^{N^{\mathrm{T}}} \frac{1}{N_k^{\mathrm{M}}} \left\| \boldsymbol{P}^k - \boldsymbol{D}^k \boldsymbol{C}^k \right\|_2^2 + \lambda \left\| \boldsymbol{C} \right\|_0$$

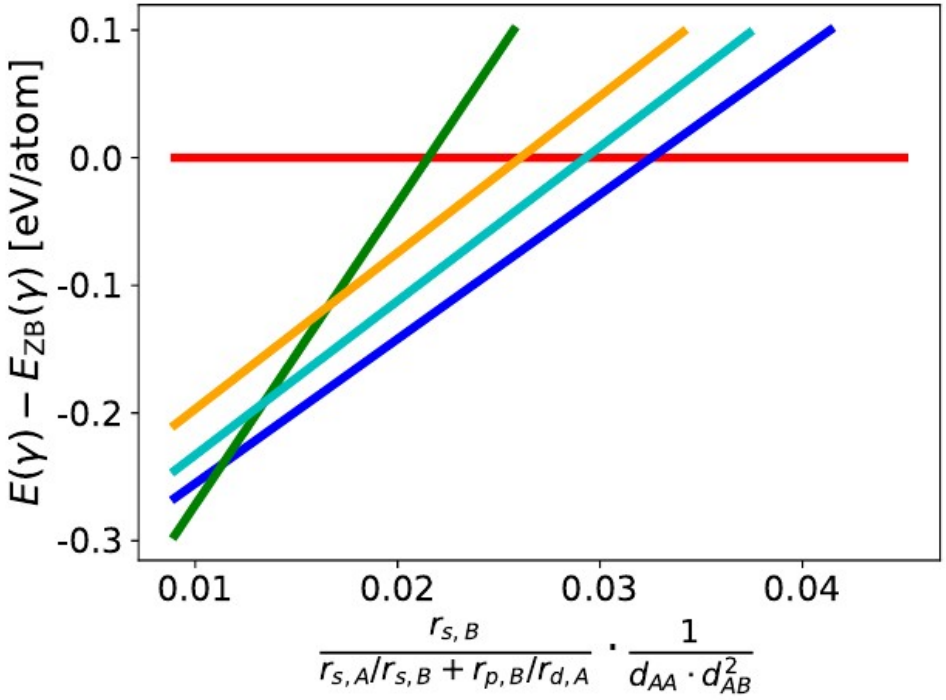
Application: multi-phase stability diagram
Properties: crystal-structure formation energies



NOMAD
NOVEL MATERIALS DISCOVERY

One descriptor to rule them all: Multi-task SISSO
Energy differences among 5 crystal structures.





Ouyang, Ahmetcik, Carbogno, Scheffler & LMG, arXiv:1901.00948

SISSO for classification

NOVEL MATERIALS DISCOVERY

$$\underset{\boldsymbol{c}\in\mathbb{R}^M}{\operatorname{argmin}}\left(\left\|\boldsymbol{P}-\boldsymbol{D}\boldsymbol{c}\right\|_2^2+\lambda\left\|\boldsymbol{c}\right\|_0\right)$$

New cost function to be minimized:
overlap of *convex* domains

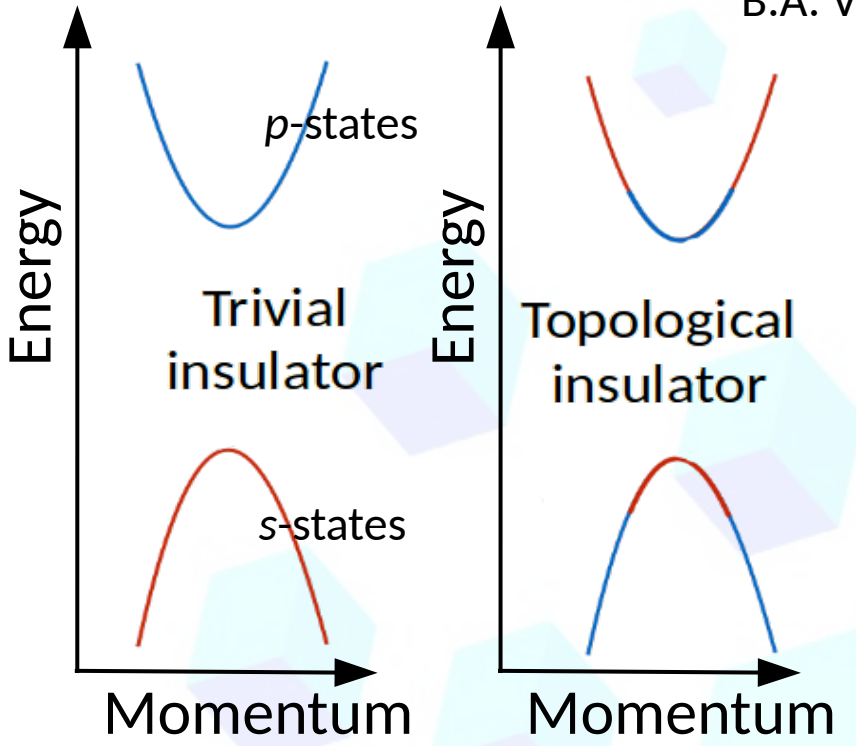
1. # points in the *convex* overlap domain
2. Area of the domain overlap
3. Distance between domains

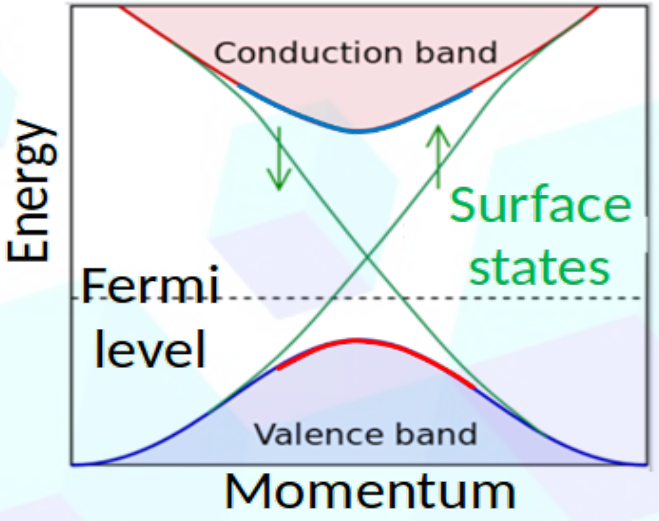
Good also for multi-categorical problems
(see A. F. Bialon *et al.*, Chem. Mater. **28**, 2550 (2016))

SISSO: predicting novel honeycomb (~2D) topological insulators

NOVEL MATERIALS DISCOVERY

B.A. Volkov, O.A. Pankratov, JETP Lett. **42**, 178 (1985)





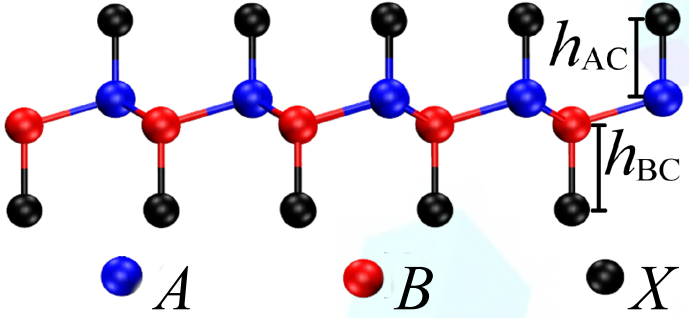
Triggered by:
spin-orbit coupling, alloying, strain, electric field

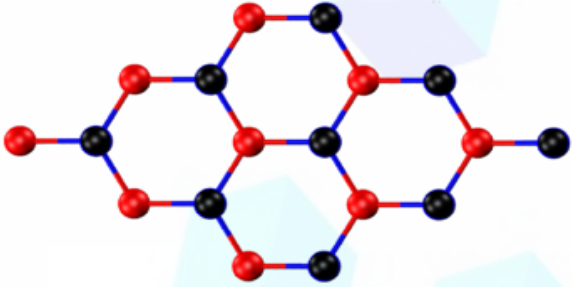
2D topological insulators:
No backscattering in edge states.
Characterized by a Z_2 topological index.

Promising materials in spintronics applications, e.g., quantum computing.

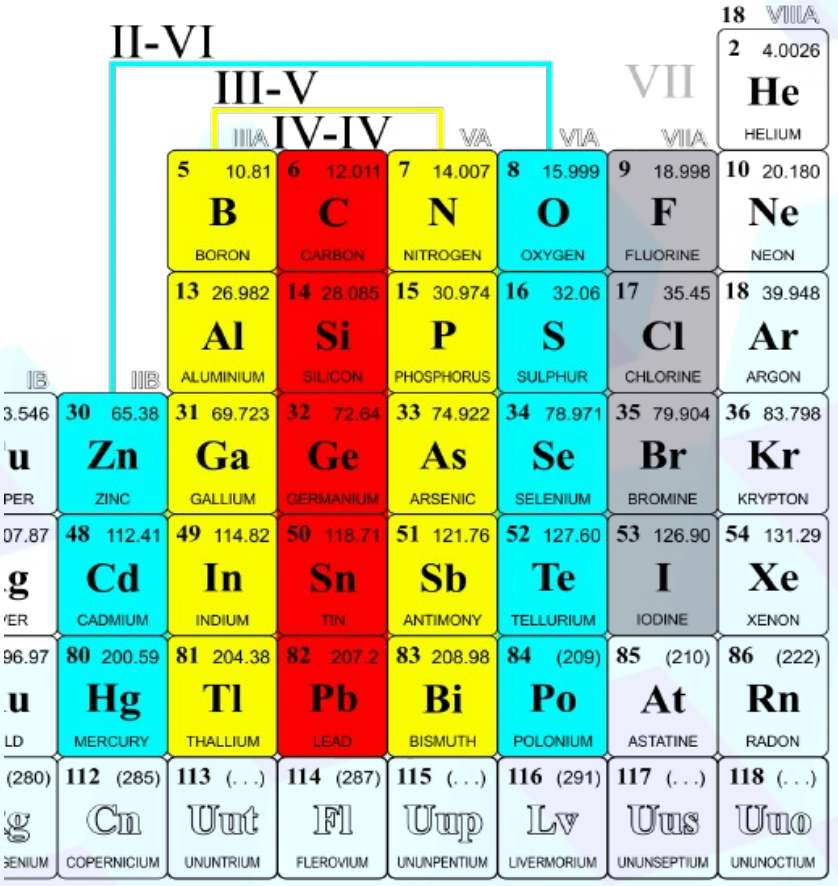
SISSO: predicting novel honeycomb (~2D) topological insulators

NOVEL MATERIALS DISCOVERY





A	II	III	IV
B	VI	V	IV
X	VII	VII	VII

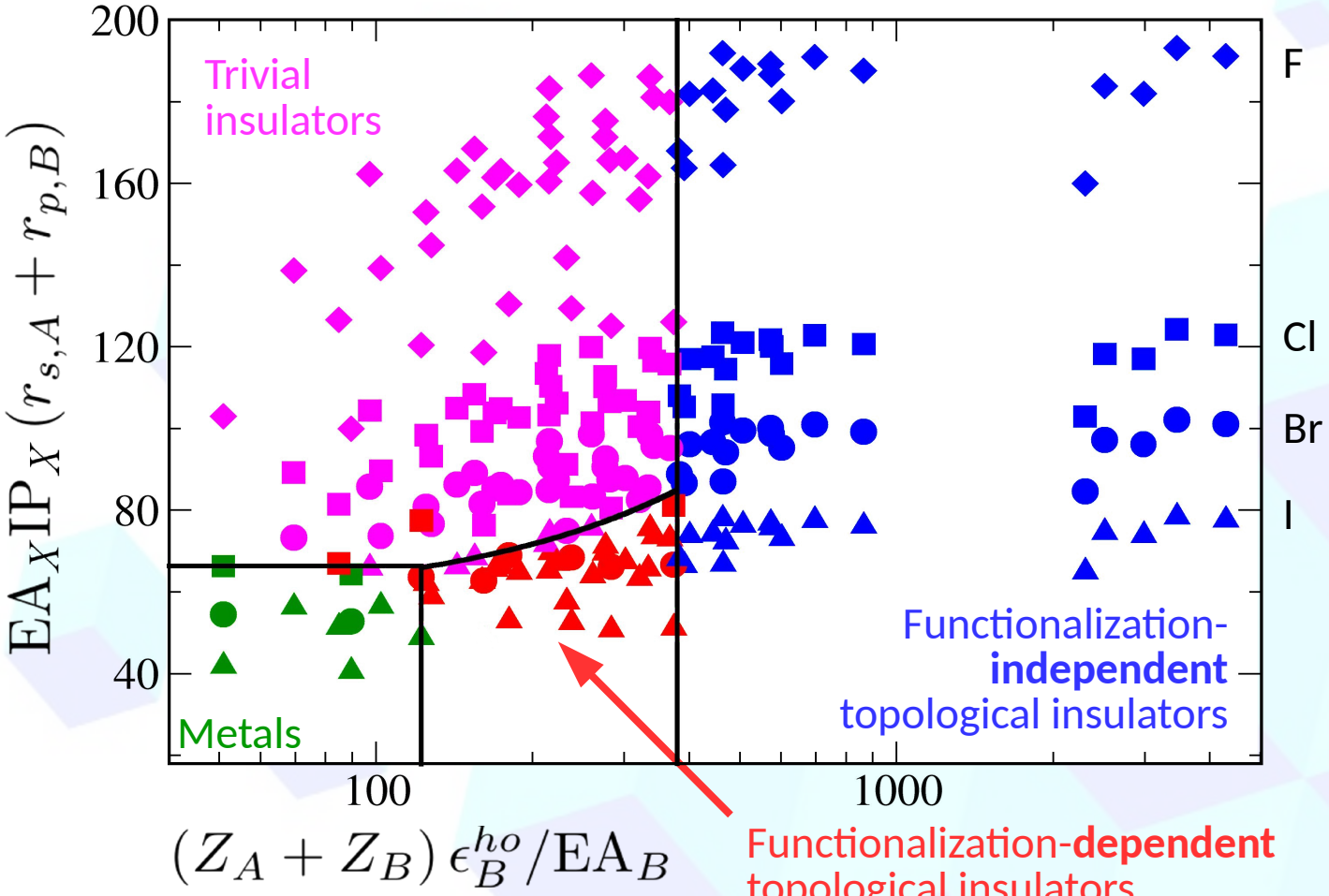


Data source: high throughput DFT (FHI-aims, Carlos Mera Acosta)

NOMAD

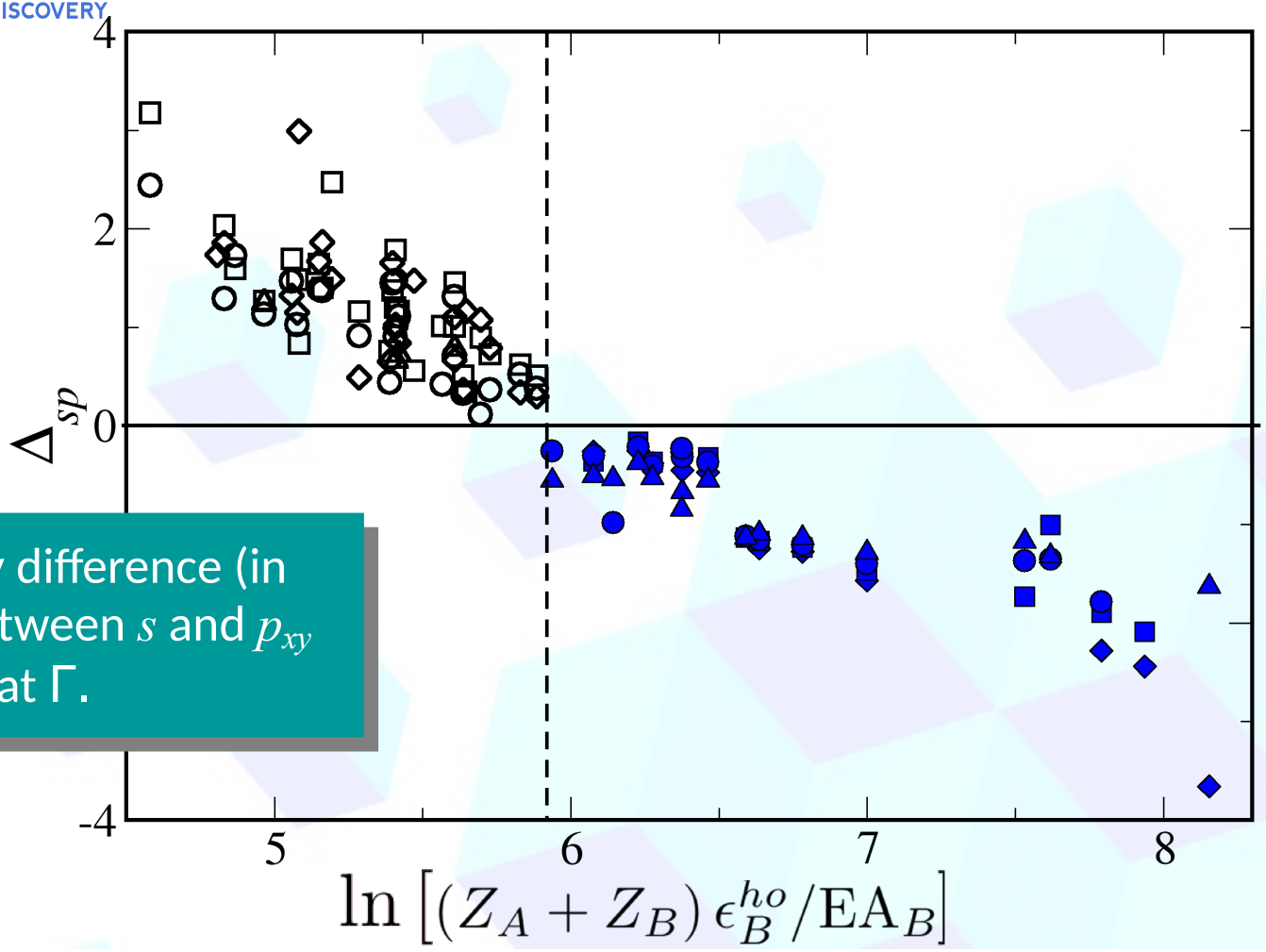
NOVEL MATERIALS DISCOVERY

SISSO: predicting novel honeycomb (~2D) topological insulators



SISSO: predicting novel honeycomb (~2D) topological insulators

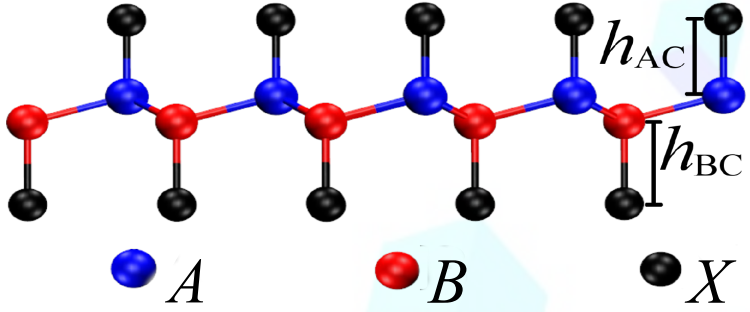
NOVEL MATERIALS DISCOVERY

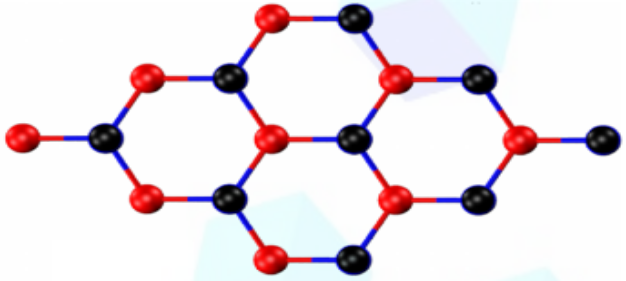


Energy difference (in eV) between s and p_{xy} states at Γ.

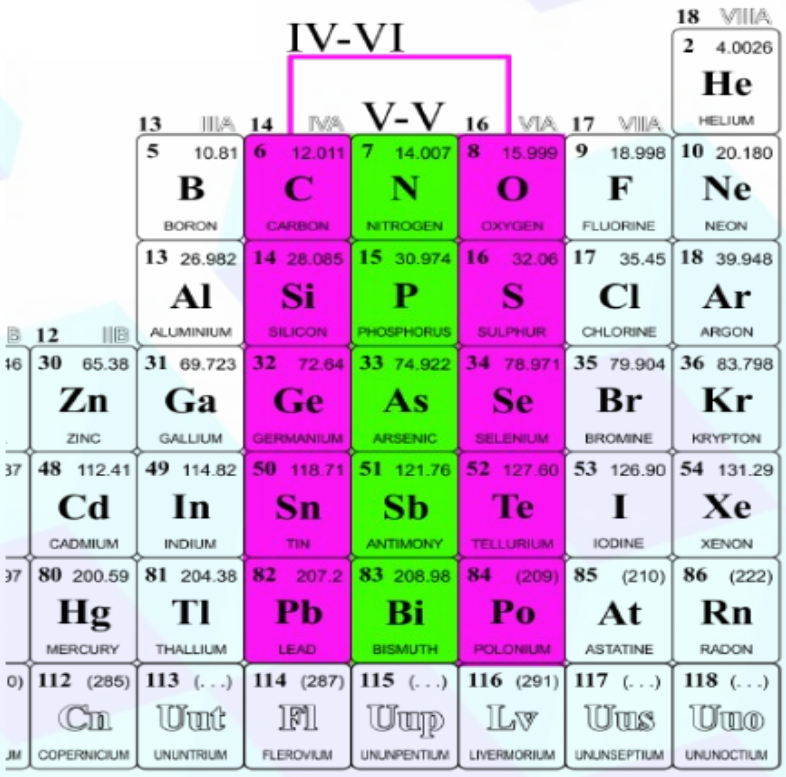
SISSO: predicting novel honeycomb (~2D) topological insulators

NOVEL MATERIALS DISCOVERY





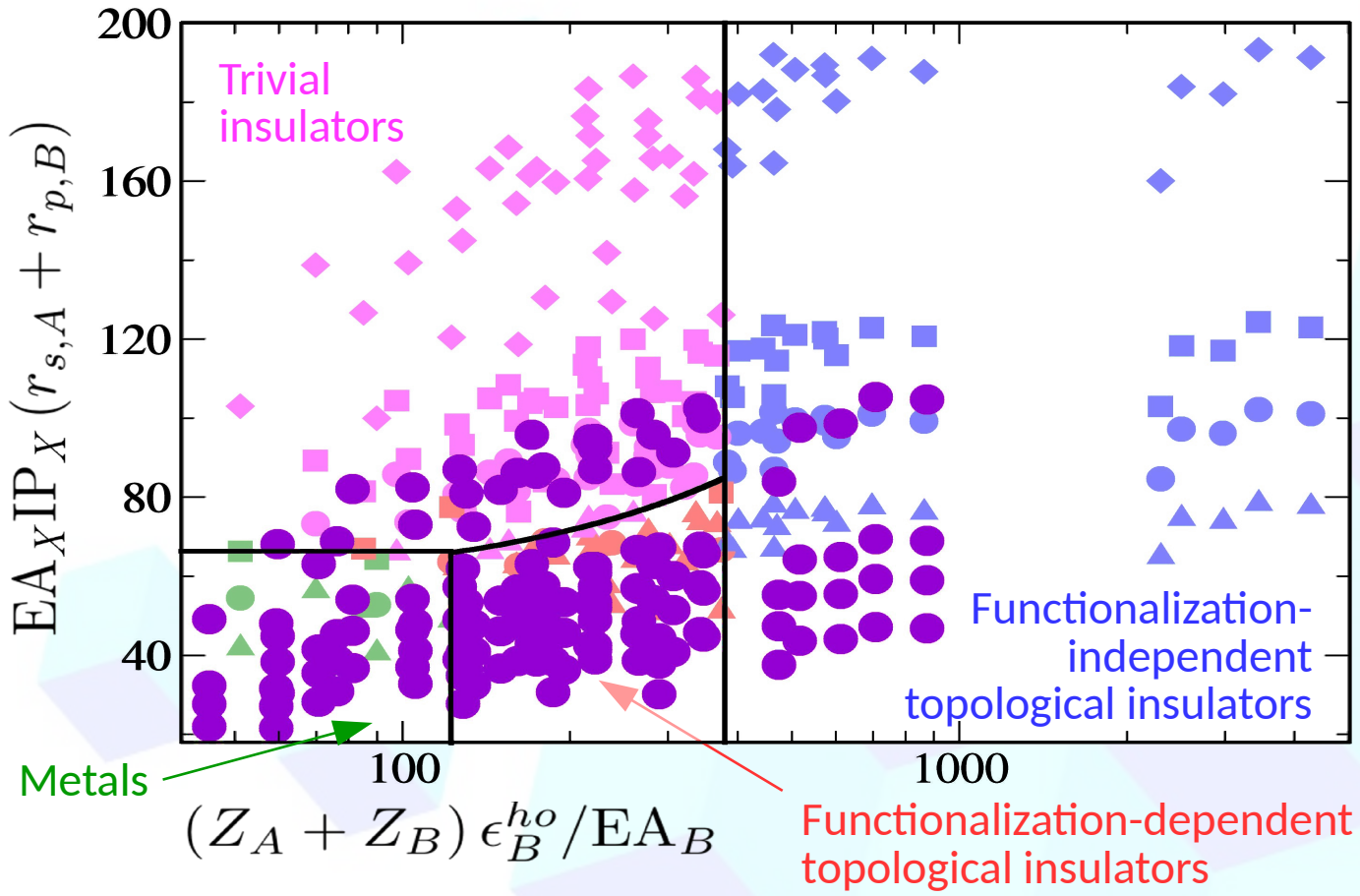
A	II	III	IV	IV V
B	VI	V	IV	VI V
X	VII	VII	VII	VI VI



Data source: high throughput DFT (FHI-aims, Carlos Mera Acosta)

SISSO: predicting novel honeycomb (~2D) topological insulators

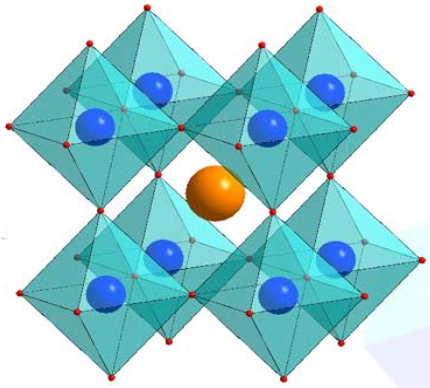
NOVEL MATERIALS DISCOVERY



Mera Acosta, Ouyang, Fazzio, Scheffler, LMG & Carbogno, arXiv: 1805.10950a

Perovskites' stability: Improving on Goldschmidt Tolerance Factor

NOVEL MATERIALS DISCOVERY



ABX_3

$$t = \frac{r_A + r_X}{\sqrt{2}(r_B + r_X)}$$ → Ionic radius

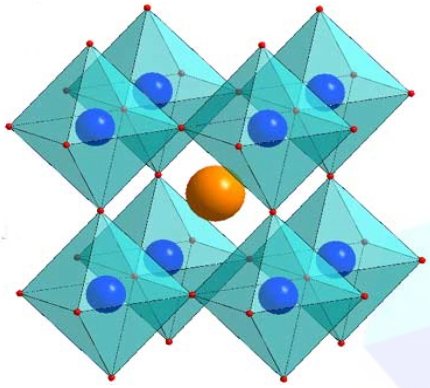
Goldschmidt* stable perovskites: accuracy 74%

t = 0.825 t = 1.059 t

Bartel, Sutton, Goldsmith, Ouyang, Musgrave, LMG &Scheffler, Sci. Adv., in press. arXiv:1801.07700

Perovskites' stability: Improving on Goldschmidt Tolerance Factor

NOVEL MATERIALS DISCOVERY



ABX_3

$$t = \frac{r_A + r_X}{\sqrt{2}(r_B + r_X)}$$

r_A, r_X: Ionic radius

$$\tau = \frac{r_X}{r_B} - n_A \left(n_A - \frac{r_A/r_B}{\ln(r_A/r_B)} \right)$$

n_A: Oxidation state

$\frac{r_X}{r_B}$: 1 / μ = Octahedral factor

Goldschmidt* stable perovskites: accuracy 74%

t = 0.825 t = 1.059 t

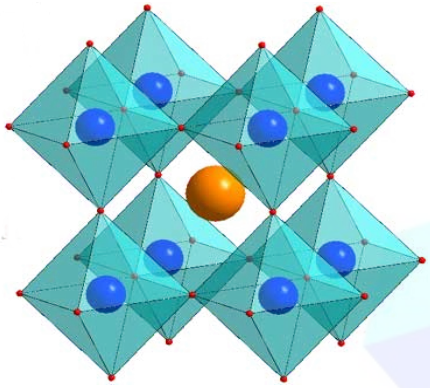
Our stable perovskites: accuracy 92%

τ = 4.18 τ

Bartel, Sutton, Goldsmith, Ouyang, Musgrave, LMG &Scheffler, Sci. Adv., in press. arXiv:1801.07700

Perovskites' stability: Improving on Goldschmidt Tolerance Factor

NOVEL MATERIALS DISCOVERY



ABX_3

$$t = \frac{r_A + r_X}{\sqrt{2}(r_B + r_X)}$$

r_A, r_X, r_B → Ionic radius

n_A → Oxidation state

$$\tau = \frac{r_X}{r_B} - n_A\left(n_A - \frac{r_A/r_B}{\ln(r_A/r_B)}\right)$$

$\frac{r_X}{r_B}$ → 1 / μ = Octahedral factor

Goldschmidt* stable perovskites: accuracy 74%

t = 0.825 t = 1.059 t

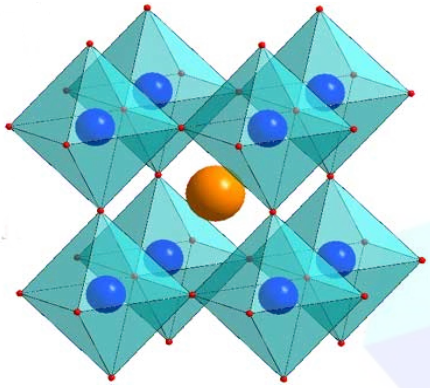
Our stable perovskites: accuracy 99%

τ = 3.31 τ = 4.18 τ = 5.92 τ

Bartel, Sutton, Goldsmith, Ouyang, Musgrave, LMG &Scheffler, Sci. Adv., in press. arXiv:1801.07700

Perovskites' stability: Improving on Goldschmidt Tolerance Factor

NOVEL MATERIALS DISCOVERY



ABX_3

$$t = \frac{r_A + r_X}{\sqrt{2}(r_B + r_X)}$$ → Ionic radius

$$\tau = \frac{r_X}{r_B} - n_A \left(n_A - \frac{r_A/r_B}{\ln(r_A/r_B)} \right)$$ → Oxidation state

$\frac{r_X}{r_B}$ → 1 / μ = Octahedral factor

Goldschmidt* stable perovskites: accuracy 74%

t = 0.825 t = 1.059 t

Our stable perovskites: accuracy 100%

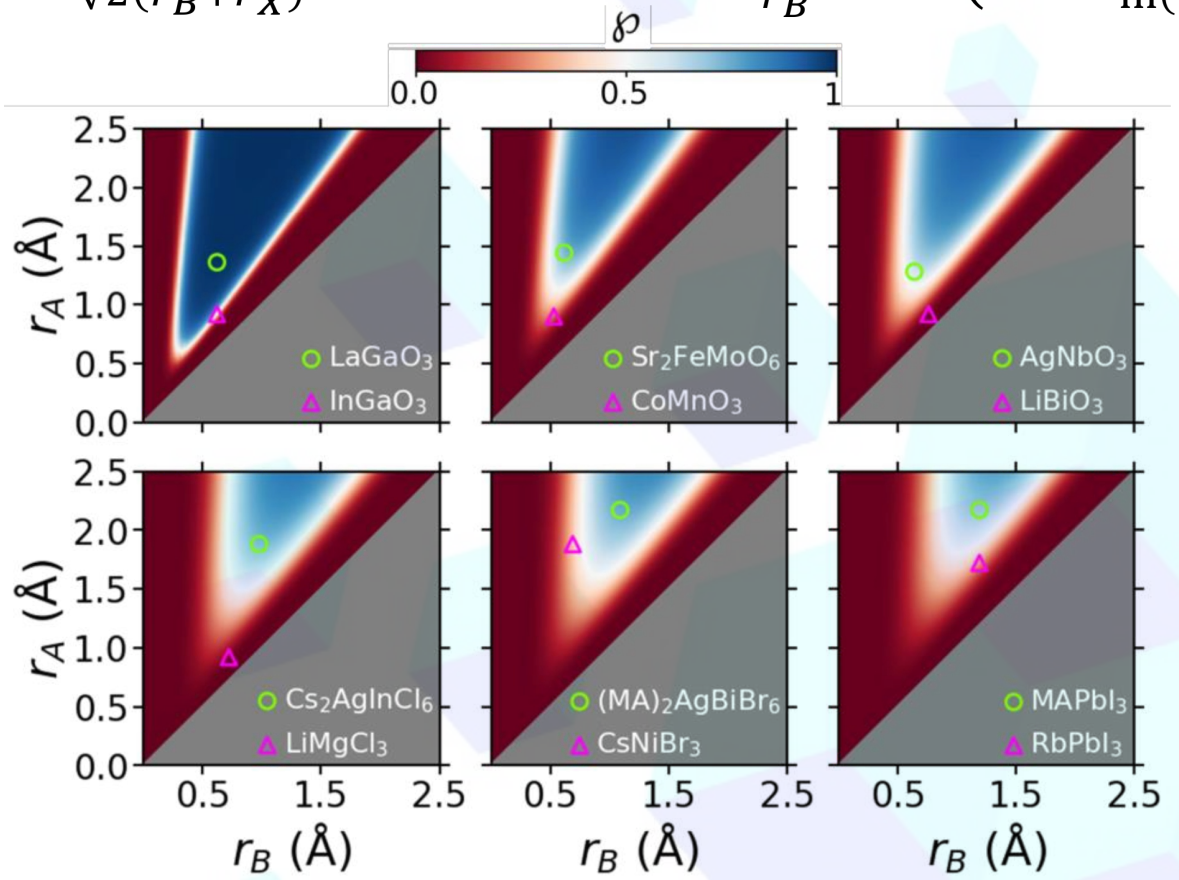
τ = 3.31 τ = 4.18 τ = 5.92 τ = 12.08 τ

Bartel, Sutton, Goldsmith, Ouyang, Musgrave, LMG &Scheffler, Sci. Adv., in press. arXiv:1801.07700

Perovskites' stability: Improving on Goldschmidt Tolerance Factor

NOVEL MATERIALS DISCOVERY

$$t = \frac{r_A + r_X}{\sqrt{2}(r_B + r_X)} \longrightarrow \tau = \frac{r_X}{r_B} - n_A\left(n_A - \frac{r_A/r_B}{\ln(r_A/r_B)}\right)$$



n_A

$X = O^{2-}$

$n_A = 1^+$

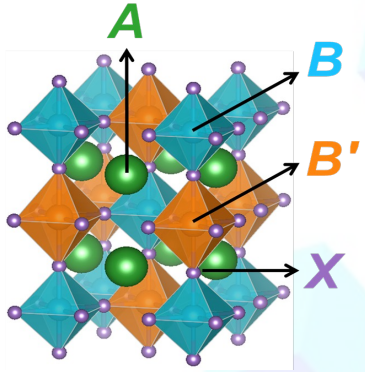
X

Perovskites' stability: Improving on Goldschmidt Tolerance Factor

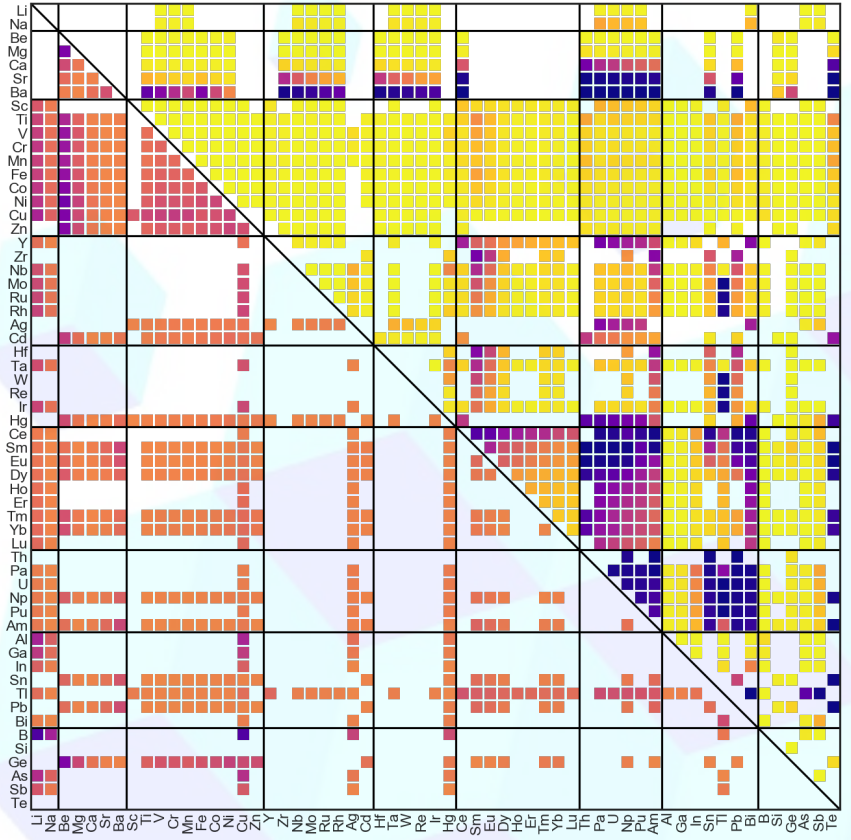
NOVEL MATERIALS DISCOVERY

$$t = \frac{r_A + r_X}{\sqrt{2}(r_B + r_X)} \longrightarrow \tau = \frac{r_X}{r_B} - n_A\left(n_A - \frac{r_A/r_B}{\ln(r_A/r_B)}\right)$$

Double perovskites

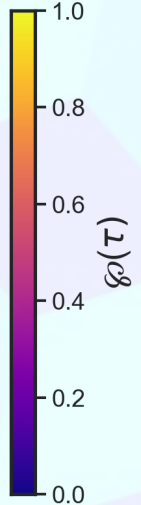


$r_B \rightarrow (r_B + r_{B'})/2$



$Cs_2BB'Cl_6$

$La_2BB'O_6$



… and more

Continuous property

Adsorption energy and OCO angle of adsorbed CO_2 on metal-oxide surfaces

Adsorption energy of O on metal-oxide surfaces

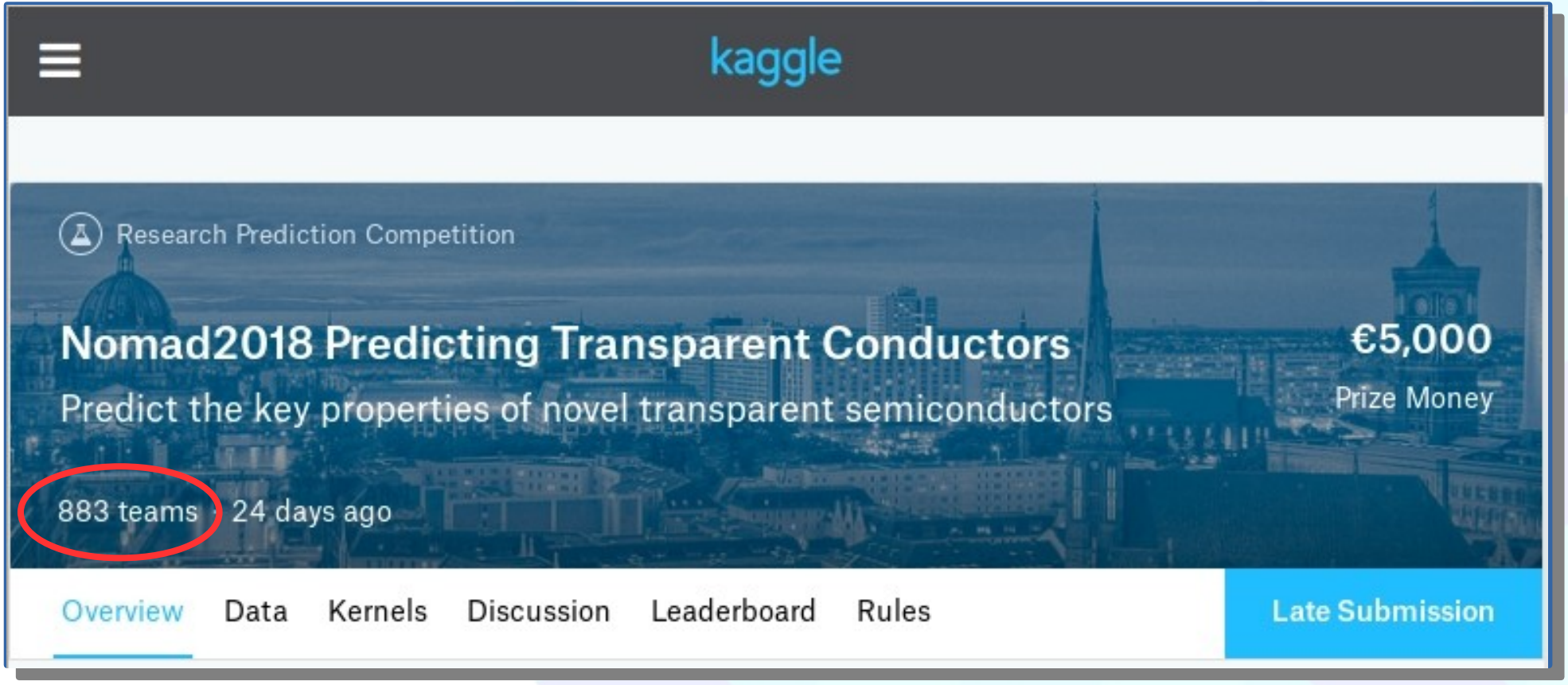
Adsorption energy of metal atoms on metal-alloys surfaces
Features: atoms (of the surface) and pristine surface

Classification

Tetradymite 5-component 3d topological insulators (vs trivial insulators)

Crowd sourcing data-driven materials science: nomad-coe.eu/outreach/nomad-kaggle-2018

NOVEL MATERIALS DISCOVERY



Crowd sourcing data-driven materials science: nomad-coe.eu/outreach/nomad-kaggle-2018

NOVEL MATERIALS DISCOVERY

- Materials class: $Al_xGa_yIn_zO_3$, $x + y + z = 3$, in 6 crystal prototypes (C2/m (12), Pna21 (33), R-3c (167), p63/mmc (194), Ia-3 (206), and Fd-3m, up to 96 atoms per unit cell
- Properties: Formation energy and band gap
- **Input: *Vegard's-law* lattice vectors and coordinates**
- ~3000 data points (2400 training + 600 validation set) calculated via PBE, by FHI-aims + Christopher Sutton

Crowd sourcing data-driven materials science: nomad-coe.eu/outreach/nomad-kaggle-2018

NOVEL MATERIALS DISCOVERY

Prediction accuracy	
Formation energy:	~20 meV/cation
Band gap:	~80 meV

1st **Takenori Yamamoto (Tony Y.)**
Descriptor: Crystal-graph *n*-grams
ML method: Kernel ridge regression

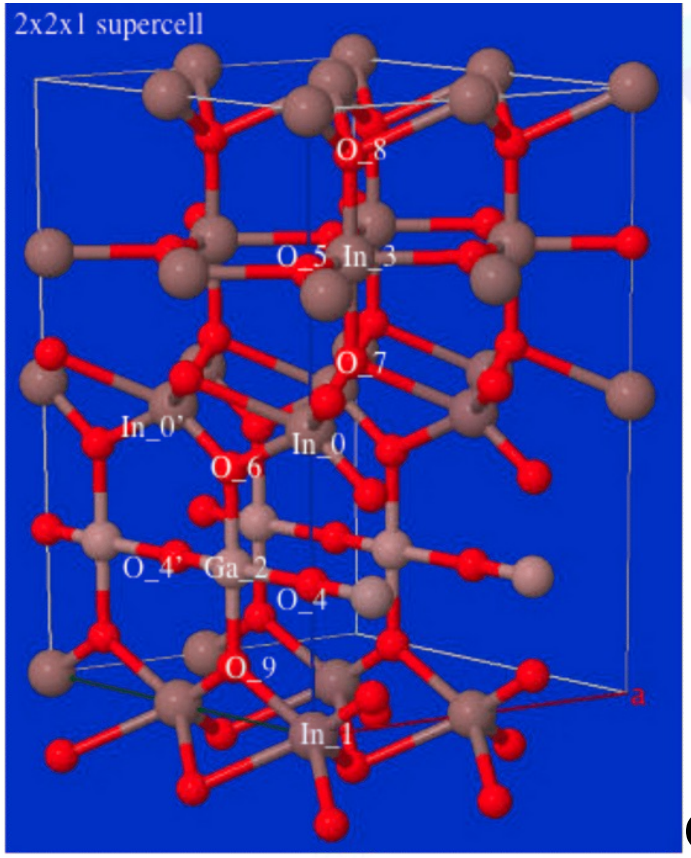
2nd **Dr. Yury Lysogorskiy**
Descriptor: 7000 average and local-order scalar features
ML method: Gradient boosting regression trees

3rd **Lars Blumenthal**
Descriptor: based on Smooth Overlap of Atomic Positions (SOAP)
ML method: Neural network

(Average) SOAP: S De, AP Bartók, G Csányi, M Ceriotti Phys. Chem. Chem. Phys. 18, 13754 (2016)
Sutton, LMG, Yamamoto, Lysogorskiy, Blumenthal, Hammerschmidt, Golebiowski, Liu, Ziletti, Scheffler arXiv:1812.00085

Crowd sourcing data-driven materials science: nomad-coe.eu/outreach/nomad-kaggle-2018

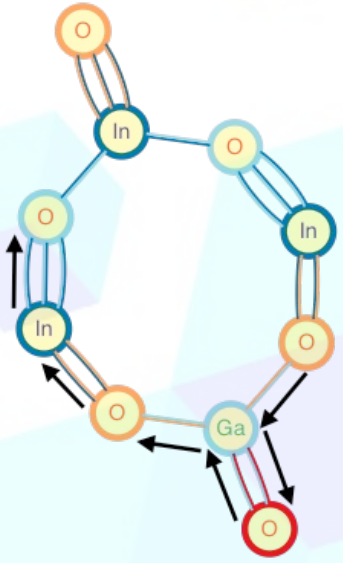
NOVEL MATERIALS DISCOVERY



$GaIn_3O_6$

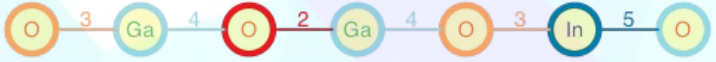
n-grams

Crystal Graph



Coordination Number

Path Graph

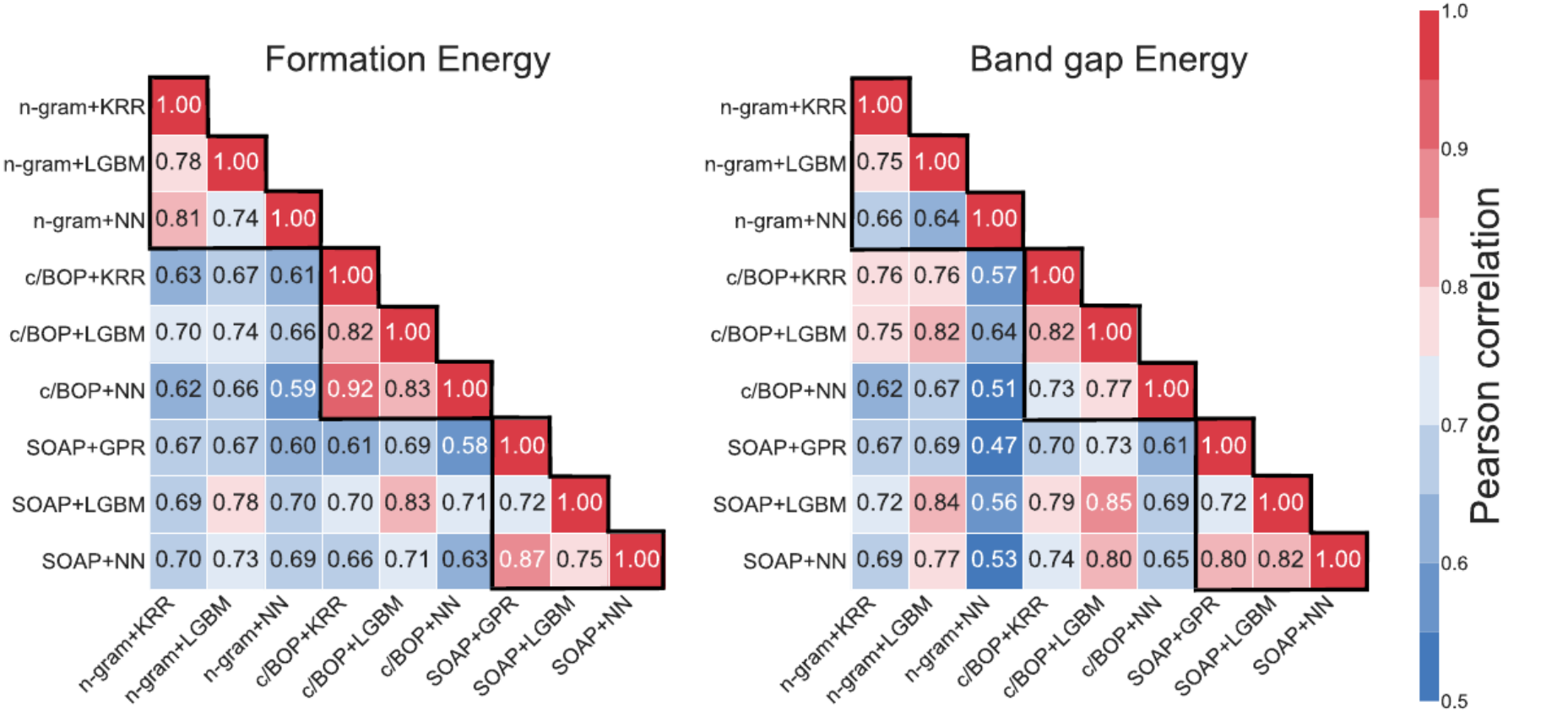


Sequence of N-gram Items

Sutton, LMG, Yamamoto, Lysogorskiy, Blumenthal, Hammerschmidt, Golebiowski, Liu, Ziletti, Scheffler arXiv:1812.00085

Crowd sourcing data-driven materials science: The importance of the descriptor

NOVEL MATERIALS DISCOVERY



Sutton, LMG, Yamamoto, Lysogorskiy, Blumenthal, Hammerschmidt, Golebiowski, Liu, Ziletti, Scheffler arXiv:1812.00085

The NOMAD Laboratory https://nomad-coe.eu

NOVEL MATERIALS DISCOVERY

The NOMAD Laboratory
A European Centre of Excellence

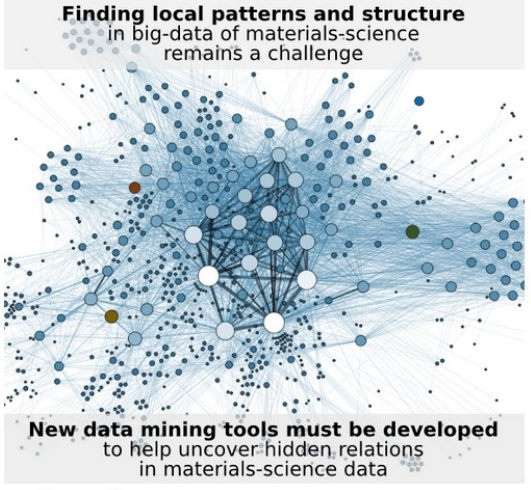
The Novel Materials Discovery (NOMAD) Laboratory maintains the largest Repository, for input and output files of all important computational materials science codes.

From its open-access data, it builds several ***Big-Data Services*** helping to advance materials science and engineering.

To learn more, click on the buttons above. You can also watch our 3-minute summary on the ***NOMAD Laboratory CoE*** at YouTube (or at YOUKU in China).

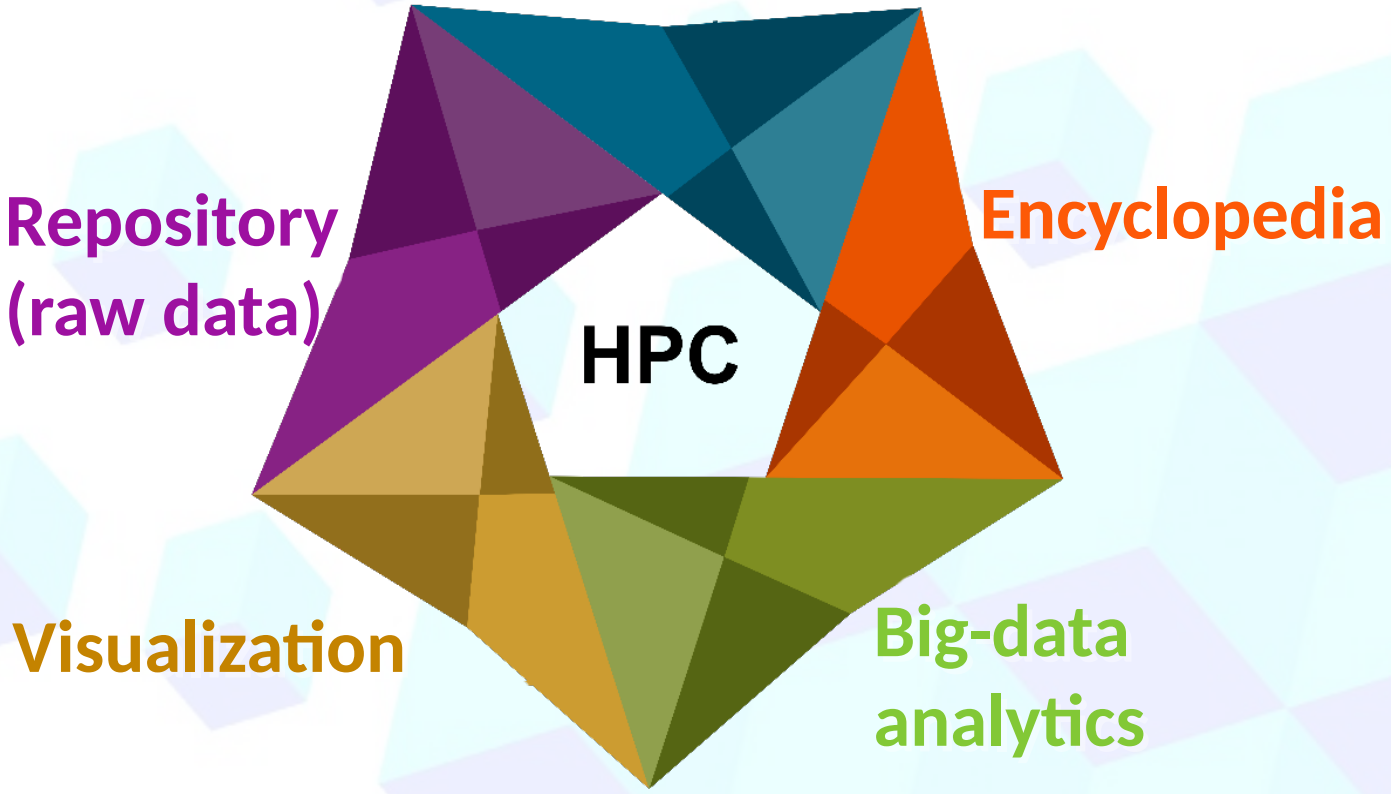
NOMAD Scope and Overview

NOMAD Success Stories



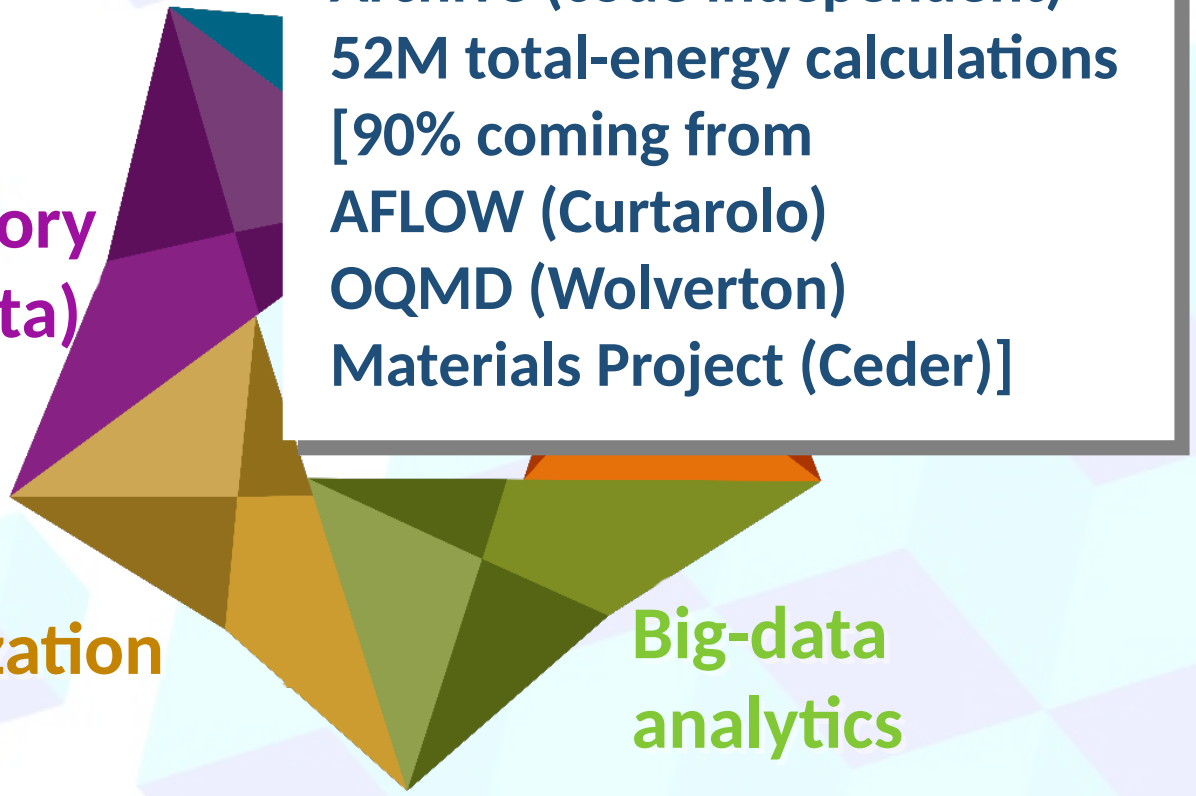
The NOMAD Laboratory https://nomad-coe.eu

NOVEL MATERIALS DISCOVERY



The NOMAD Laboratory https://nomad-coe.eu

NOVEL MATERIALS DISCOVERY



Archive (code independent)
52M total-energy calculations
[90% coming from
AFLOW (Curtarolo)
OQMD (Wolverton)
Materials Project (Ceder)]

The NOMAD Laboratory https://nomad-coe.eu

NOVEL MATERIALS DISCOVERY

The NOMAD Laboratory
A European Centre of Excellence

HOME PROJECT INDUSTRY TEAM RELATED PROJECTS NEWS PRESS KIT CONTACT US

Enter Search...

NOMAD REPOSITORY | THE ARCHIVE | ENCYCLOPEDIA | BIG-DATA ANALYTICS | ADVANCED GRAPHICS | HPC INFRASTRUCTURE | OUTREACH

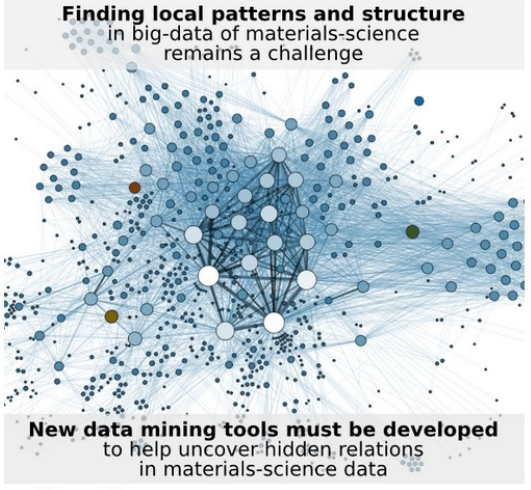
The Novel Materials Discovery (NOMAD) Laboratory maintains the largest Repository, for input and output files of all important computational materials science codes.

From its open-access data, it builds several ***Big-Data Services*** helping to advance materials science and engineering.

To learn more, click on the buttons above. You can also watch our 3-minute summary on the ***NOMAD Laboratory CoE*** at YouTube (or at YOUKU in China).

NOMAD Scope and Overview

NOMAD Success Stories



The NOMAD Laboratory https://nomad-coe.eu

NOVEL MATERIALS DISCOVERY

Materials property prediction

Predicting the metal-insulator classification of elements and binary systems

Prediction of topological quantum phase transitions

NOMAD Analytics Toolkit

– Prediction of topological quantum phase transitions

created by: Carlos Mera Acosta[1,2], Emre Ahmetcik [1], Adalberto Fazzio[2], Luca Ghiringhelli[1], and Matthias Scheffler[1]

[1] Fritz Haber Institute of the Max Planck Society, Faradayweg 4-6, D-14195 Berlin, Germany
[2] Brazilian Nanotechnology National Laboratory, Campinas, SP, Brazil

v1.0.0

[Last updated: November 9, 2017]

Back to Analytics Home | Back to NOMAD CoE Home

Acknowledgements

NOVEL MATERIALS DISCOVERY

SISSO and Multi-Task SISSO
Runhai Ouyang, Emre Ahmetcik, Stefano Curtarolo, Christian Carbogno

Topological Insulators
Carlos Mera Acosta, Adalberto Fazzio, Runhai Ouyang, Christian Carbogno

Perovskites
Christopher J. Bartel, Christopher Sutton, Bryan R. Goldsmith, Runhai Ouyang, Charles B. Musgrave

NOMAD-Kaggle competition
Christopher Sutton, Angelo Ziletti, Xiangyue Liu

NOMAD PIs
Angel Rubio (MPSD Hamburg), Risto Nieminen (Aalto Univ. Helsinki), Francesc Illas (Univ. Barcelona), Daan Frenkel (Univ. Cambridge), Claudia Draxl (HU Berlin), Alessandro De Vita (Kings College London), Kristian Thygesen (DTU Lyngby); Kimmo Koski (CSC Helsinki), Stefan Heinzel (MPSCD Garching), Jose Maria Cela (BSC Barcelona), Dieter Kranzlmüller (LRZ Munich); Ciaran Clissman (Pintail Dublin)

All the above
Matthias Scheffler

NOMAD has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 676580.

BERLIN BIG DATA CENTER